

WHAT'S NEW!

SEARCH

[Keyword](#)
[Site Map](#)
[Test Locator](#)

LIBRARY

[Reports](#)
[Newsletters](#)
[Assessments](#)
[Rubrics](#)
[Products](#)
[Overheads](#)

SERVICES

[Ask the Expert](#)

MEDIA CONTACTS

PARENTS' PAGE

SPECIAL INTERESTS

[Related Links](#)
[Glossary](#)
[CRESST &
LAUSD](#)
[Rubrics](#)



Glossary

CRESST Assessment Glossary

For purposes of clarity we have cross-referenced and defined important assessment terms that occur within each definition. Cross- references are in bold.

Accommodations and Adaptations. Modifications in the way assessments are designed or administered so that **students with disabilities (SWD)** and limited English proficient students can be included in the assessment. Assessment accommodations or adaptations might include Braille forms for blind students or tests in native languages for students whose primary language is other than English.

Alignment The process of linking content and performance standards to assessment, instruction, and learning in classrooms. One typical alignment strategy is the step-by-step development of (a) content standards, (b) performance standards, (c) assessments, and (d) instruction for classroom learning. Ideally, each step is informed by the previous step or steps, and the sequential process is represented as follows:

Content Standards - Performance Standards - Assessments - Instruction for Learning

In practice, the steps of the alignment process will overlap. The crucial question is whether classroom teaching and learning activities support the standards and assessments. System alignment also includes the link between other school, district, and state resources. Alignment supports the goals of the standards, i.e., whether professional development priorities and instructional materials are linked to what is necessary to achieve the standards.

Alternative Assessment (also authentic or performance assessment). An assessment that requires students to generate a response to a question rather than choose from a set of responses provided to them. Exhibitions, investigations, demonstrations, written or oral responses, journals, and **portfolios** are examples of the assessment alternatives we think of when we use the term "alternative assessment." Ideally, alternative assessment requires students to actively accomplish complex and significant tasks, while bringing to bear prior

JOBS AT CRESST



HOME

knowledge, recent learning, and relevant skills to solve realistic or authentic problems. Alternative assessments are usually one key element of an **assessment system**.

Analytic Scoring Evaluating student work across multiple **dimensions** of performance rather than from an overall impression (**holistic scoring**). In analytic scoring, individual scores for each dimension are scored and reported. For example, analytic scoring of a history essay might include scores of the following dimensions: use of prior knowledge, application of principles, use of original source material to support a point of view, and composition. An overall impression of quality may be included in analytic scoring.

Anchor(s). A sample of student work that exemplifies a specific level of performance. Raters use anchors to score student work, usually comparing the student performance to the anchor. For example, if student work was being scored on a scale of 1-5, there would typically be anchors (previously scored student work), exemplifying each point on the scale.

Assessment. The process of gathering, describing, or quantifying information about performance.

Assessment System. The combination of multiple assessments into a comprehensive reporting format that produces comprehensive, credible, dependable information upon which important decisions can be made about students, schools, districts, or states. An assessment system may consist of a **norm-referenced** or **criterion-referenced assessment**, an **alternative assessment system**, and **classroom assessments**.

Benchmark. A detailed description of a specific level of student performance expected of students at particular ages, grades, or development levels. Benchmarks are often represented by samples of student work. A set of benchmarks can be used as "checkpoints" to monitor progress toward meeting performance goals within and across grade levels, i.e., benchmarks for expected mathematics capabilities at Grades 3, 7, ten, 10 graduation.

Classroom Assessment. An assessment developed, administered, and scored by a teacher or set of teachers with the purpose of evaluating individual or classroom student performance on a topic. Classroom assessments may be aligned into an **assessment system** that includes **alternative assessments** and either a **norm-referenced** or **criterion-referenced assessment**. Ideally, the results of a classroom assessment are used to inform and influence instruction that helps students reach high standards.

Content Standards. Broadly stated expectations of what students should *know and be able to do* in particular subjects and grade levels. Content standards define for teachers, schools, students, and the community not only the expected student skills and knowledge, but what schools should teach. An example of a language arts standard is: "Fourth-grade students will be able to gather information for a report using sources such as interviews, questionnaires, computers, and library centers."

Criteria. Guidelines, rules, characteristics, or **dimensions** that are used to judge the quality of student performance. Criteria indicate what we value in student responses, products or performances. They may be holistic, analytic, general, or specific. **Scoring rubrics** are based on criteria and define what the criteria mean and how they are used.

Criterion-Referenced Assessment. An assessment where an individual's performance is compared to a specific learning objective or **performance standard** and not to the performance of other students. Criterion-referenced assessment tells us how well students are performing on specific goals or standards rather than just telling how their performance compares to a norm group of students nationally or locally. In criterion-referenced assessments, it is possible that none, or all, of the examinees will reach a particular goal or performance standard. For example: "all of the students demonstrated *proficiency* in applying concepts from astronomy, meteorology, geology, oceanography, and physics to describe the forces that shape the earth."

Dimensions. Desired knowledge or skills measured in an assessment and usually represented in a **scoring rubric**. For example, a measurement of student teamwork skills on a performance assessment might include 6 dimensions: *adaptability* (recognizing problems and responding appropriately), *coordination* (organizing team activities to complete a task on time), *decision making* (using available information to make decisions), *interpersonal* (interacting cooperatively with other team members), *leadership* (providing direction for the team), and *communication* (clearly and accurately exchanging information between team members).

Equity. Equity is the concern for fairness, i.e., that assessments are free from bias or favoritism. An assessment that is fair enables *all* children to show what they can do. At minimum, all assessments should be reviewed for (a) stereotypes, (b) situations that may favor one culture over another, (c) excessive language demands that prevent some students from showing their knowledge, and (d) the assessment's potential to include students with disabilities or limited English proficiency.

Evaluation. When used for most educational settings, evaluation means to measure, compare, and judge the quality of student work, schools, or a specific educational program.

Holistic Scoring. Evaluating student work in which the score is based on an overall impression of student performance rather than multiple dimensions of performance (**analytic scoring**).

Item. An individual question or exercise in an assessment or evaluative instrument.

Judge. See rater.

Norm-Referenced Assessment. An assessment where student performance or performances are compared to a larger group. Usually the larger group or "norm group" is a national sample representing a wide and diverse cross-section of students. Students, schools, districts, and even states are compared or rank-ordered in relation to the norm group. The purpose of a norm-referenced assessment is usually to sort students and not to measure achievement towards some criterion of performance.

On-Demand Assessment. An assessment that takes place at a predetermined time and place, usually under uniform conditions for all students being assessed. The SAT, district and state tests, and most in-class unit tests and final exams are examples of on-demand assessments.

Opportunity to Learn. To provide students with the teachers, materials, facilities, and instructional experiences that will enable them to achieve high standards. Opportunity to learn (OTL) is what takes place in classrooms that enables students to acquire the knowledge and skills that are expected. OTL can include *what* is taught, *how* it is taught, by *whom*, and with *what* resources.

Performance Assessment. See alternative assessment.

Performance Standards. Explicit definitions of what students must do to demonstrate proficiency at a specific level on the **content standards**. For example, the performance level "exceptional achievement" on a **dimension** "communication of ideas" is reached when the student *examines the problem from several different positions and provides adequate evidence to support each position*.

Portfolio Assessment. A portfolio is collection of work, usually drawn from students' classroom work. A portfolio becomes a portfolio assessment when (1) the assessment purpose is defined; (2) criteria or methods are made clear for determining *what* is put into the portfolio, *by whom*, and *when*; and (3) criteria for assessing either the collection or individual pieces of work are identified and used to make judgments about performance. Portfolios can be designed to assess student progress, effort, and/or achievement, and encourage students to reflect on their learning.

Rater. A person who evaluates or judges student performance on an assessment against specific criteria.

Rater Training. The process of educating **raters** to evaluate student work and produce dependable scores. Typically, this process uses **anchors** to acquaint raters with criteria and scoring rubrics. Open discussions between raters and the trainer help to clarify **scoring criteria** and **performance standards**, and provide opportunities for raters to practice applying the rubric to student work. Rater training often includes an assessment of rater reliability, that raters must pass in order to score actual student work.

Reliability. The degree to which the results of an assessment are dependable and consistently measure particular student knowledge and/or skills. Reliability is an indication of the consistency of scores across **raters**, over time, or across different **tasks** or **items** that measure the same thing. Thus, reliability may be expressed as (a) the relationship between test items intended to measure the same skill or knowledge (item reliability), (b) the relationship between two administrations of the same test to the same student or students (test/retest reliability), or (c) the degree of agreement between two or more raters (rater reliability). An unreliable assessment cannot be valid.

Scale. Values given to student performance. Scales may be applied to individual items or performances, for example, *checklists*, i.e., yes or no; *numerical*, i.e., 1-6; or *descriptive*, i.e., the student presented multiple points of view to support her essay. Scaled scores occur when participants' responses to any number of items are combined and used to establish and place students on a single scale of performance.

Standardization. A consistent set of procedures for designing, administering, and scoring an assessment. The purpose of standardization is to assure that all students are assessed under the same conditions so that their scores have the same meaning and are not influenced by differing conditions. Standardized procedures are very important when scores will be used to compare individuals or groups.

Scorer. See **Rater**.

Standards. The broadest of a family of terms referring to statements of expectations for student learning, including **content standards**, **performance standards**, and **benchmarks**.

Standards-Based Reform. A program of school improvement involving setting high standards for all students and a process for adapting instruction and assessment to make sure all students can achieve the standards.

Students With Disabilities (SWD). A broadly defined group of students with physical and/or mental impairments such as blindness or learning disabilities that might make it more difficult for them to do well on assessments without **accommodations** or **adaptations**.

Task. An activity, exercise, or question requiring students to solve a specific problem or demonstrate knowledge of specific topics or processes.

Validity. The extent to which an assessment measures what it is supposed to measure and the extent to which inferences and actions made on the basis of test scores are appropriate and accurate. For example, if a student performs well on a reading test, how confident are we that that student is a good reader? A valid standards-based assessment is aligned with the **standards** intended to be measured, provides an accurate and reliable estimate of students' performance relative to the standard, and is fair. An assessment cannot be valid if it is not reliable.

CRESST is affiliated with the [Graduate School of Education & Information Studies](#) at UCLA.

CRESST/UCLA
301 GSE&IS, Mailbox 951522
300 Charles E. Young Drive North
Los Angeles, CA 90095-1522
Tel: 310-206-1532 Fax: 310-825-3883

CONTACT the Webmaster (webmaster@cse.ucla.edu) concerning this server.

This page last modified March 9, 1999.